Spatio-structural Symbol Description with Statistical Feature Add-on

Santosh K.C., Bart Lamiroy, Laurent Wendling

Spatio-structural Symbol Description with Statistical Feature Add-on

K.C. Santosh[†] and Bart Lamiroy[♯] and Laurent Wendling[‡]

[†]INRIA Nancy Grand Est, [♯]Université de Lorraine – INPL
LORIA Campus Scientifique, BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex, France
[‡]LIPADE, Université Paris Descartes, 75270 Paris Cedex 06, France
Santosh.KC, Bart.Lamiroy@loria.fr, Laurent.Wendling@parisdescartes.fr

Abstract. In this paper, we present a method for symbol description based on both spatio-structural and statistical features computed on elementary visual parts, called 'vocabulary'. This extracted vocabulary is grouped by type (e.g., *circle*, *corner*) and serves as a basis for an attributed relational graph where spatial relational descriptors formalise the links between the vertices, formed by these types, labelled with global shape descriptors. The obtained attributed relational graph description has interesting properties that allows it to be used efficiently for recognising structure and by comparing its attribute signatures. The method is experimentally validated in the context of electrical symbol recognition from wiring diagrams.

1 Introduction

Graphics recognition has an extremely rich state-of-the-art literature in symbol recognition and localisation. However, most methods are particularly suited for isolated line symbols, not for composed symbols connected to a complex environment [1,2]. Considering the problem of symbol localisation in real documents, composed of individual parts and constrained by spatial relations for instance, one needs to be able to extract visual parts, characterise their shape description and formalise the possible links that exist between them. This integration of spatial relations and shape description of the extracted visual parts is going to be the core of this paper. The method is very much inspired by a real world industrial problem [3,4,5]. Fig. 1 shows a few samples of the data related to it.

Global signal based descriptors [6] present a number of inconvenients in our context. They difficultly accommodate with connected or composite symbols and they are generally not well adapted for capturing small detail changes. In statistical approaches, signatures are simple with low computational cost. They are, unfortunately, primarily designed for applications where line symbols are isolated [7]. Furthermore, discriminative power and robustness in general applications usually require optimal selection of features [8] or the fusion of different classifiers [9]. Besides global signal based descriptors, another idea is to decompose the symbols into either vector based primitives like *points*, *lines* and *arcs* or into meaningful parts like *circles*, *triangles* and *rectangles*. These primitives are

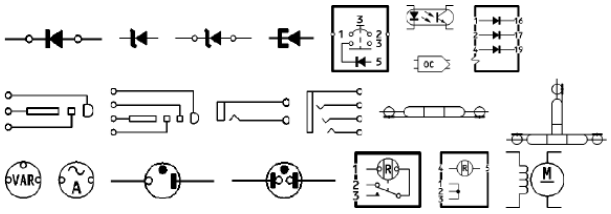


Fig. 1. A few symbols in FRESH dataset containing both linear as well as composite symbols.

then used in structural descriptors like attributed relational graphs (ARG) [10], region adjacency graphs (RAG) [11], as well as deformable templates [12]. In addition to the common drawback related to stability issues coming from segmentation and, error-prone raster-to-vector conversion, variability of the size of the underlying graph structures leads to computational complexity in matching. Structural approaches however, provide a powerful relational representation, conveying how parts are connected to each other, and are usually considered to be preserving generality and extensibility.

In this paper, we aim to combine the best of both structural and statistical approaches, and try to avoid the shortcomings of each of them. To do so, we decompose symbols by expressing their various parts in a fixed visual vocabulary, using spatial relations, graphs and signal based descriptors to describe the whole shape. Our symbol description is explained in Section 2. Symbols can be compared by computing matching scores based on vertex and edge alignment (*cf.* Section 3). In Section 4, we validate our method and compare it with the state-of-the-art. The paper concludes in Section 5. This paper is the extension of the previously published work [13], where we have validated the use of spatial relations for symbol recognition, but which did not include the vertex signatures developed here.

2 Symbol Description

Expanding on previously published work [3,13], we use a set of well controlled elementary part detectors to define a visual *vocabulary*. In our case, they consist of: *circles, corners, loose ends* and *thick* (filled) components. More formally, we denote the type set as, $\sum_{\mathbb{T}} = \{\mathbb{T}_{thick}, \mathbb{T}_{circle}, \mathbb{T}_{corner}, \mathbb{T}_{extremity}\}$. Such visual elementary parts are extracted with the help of image treatment analysis operations [14]. While, in the general case, this vocabulary can be of any kind from any type of bag-of-features, related to what is visually pertinent in the application context under consideration, our current vocabulary is related to electrical symbols. It can be easily extended to adapt to other domains.

Rather than using the detected elements as a basis for expressing and computing spatial relations, we group them together according to their types, as shown in Fig. 2. Now, we represent whole symbol by a complete ARG as a 4-tuple $G = (V, E, F_A, F_E)$ where V is the set of vertices, $E \subseteq V \times V$ is the set of graph edges, $F_A : V \rightarrow A_V$ is a function assigning attributes to the vertices and A_V representing a set of vocabulary type set $\sum_{\mathbb{T}}$ as well as their global shape

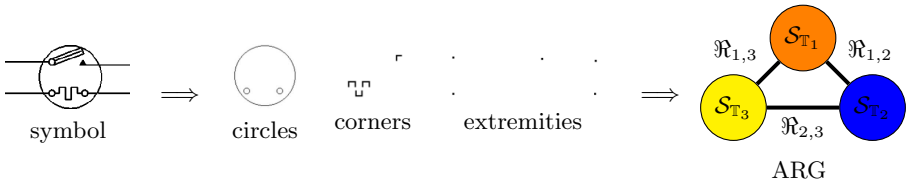


Fig. 2. ARG description for a symbol from its corresponding visual parts – an example.

signatures $\mathcal{S}$, and $F_E : E \to \Re_E$ is a function assigning labels to the edges where $\Re$ represents spatial relations of the edge E. Following the symbol in Fig. 2, the resulting graph whose attribute type set is $\{\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3\}$, can be expressed as $G = \{$

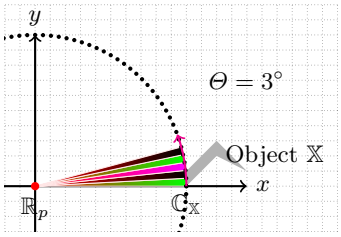
$$
\begin{aligned}
V &= \{\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3\},\\
E &= \{(\mathbb{T}_1, \mathbb{T}_2), (\mathbb{T}_1, \mathbb{T}_3), (\mathbb{T}_2, \mathbb{T}_3)\},\\
F_A &= \{((\mathbb{T}_1, \mathbb{T}_{circle}), \mathcal{S}_{\mathbb{T}_1}), ((\mathbb{T}_2, \mathbb{T}_{corner}), \mathcal{S}_{\mathbb{T}_2}), ((\mathbb{T}_3, \mathbb{T}_{extremity}), \mathcal{S}_{\mathbb{T}_3})\},\\
F_E &= \{((\mathbb{T}_1, \mathbb{T}_2), \Re_{1,2}), ((\mathbb{T}_1, \mathbb{T}_3), \Re_{1,3}), ((\mathbb{T}_2, \mathbb{T}_3), \Re_{2,3})\}\}.
\end{aligned}
$$

Since this forms a complete graph, it is obvious that there exist $r = \frac{t(t-1)}{2}$ edges for t attribute types. However, because of the use of fixed and completely labelled attributes, we can avoid the NP-hardness of the matching problem (*cf.* Section 3).

In what follows, we explain how edges are labelled with spatial relations, computed between the vertices, and how vertices are labelled with shape features. This results in symbols being represented by ARGs like the one depicted in Fig. 2.

Edge Signatures using Spatial Relations. The choice of spatial relation models depends on configuration of the studied objects. If the objects are far enough from each other, their relations can be approximated by their centres and their discretised angle [15]. Otherwise, if they are neither too far nor too close, relations can be approximated by their minimum bounding rectangle (MBR) [16] as long as the objects are regular in shape (and their respective MBR are non-overlapping). Approaches like angle histograms [17], tend to be more capable of dealing with overlapping. However, since they consider all pixels, their computational cost increases dramatically. Our work is inspired by the concept of fuzzy relations that take degrees of truth which is more natural than using standard, all-or-none relations [18].

Our model can be explained as follows. Given two shapes $\mathbb{A}$ and $\mathbb{B}$ for which we want to compute the relative position, and given a reference point $\mathbb{R}_p$, we cover the surrounding space at regular radial intervals of $\Theta = 2\pi/m$ by using a radial line. We compute this unique reference point $\mathbb{R}_p = \frac{\mathbb{C}_\mathbb{A} + \mathbb{C}_\mathbb{B}}{2}$ from the centroids $\mathbb{C}_\mathbb{A}$ and $\mathbb{C}_\mathbb{B}$ of the shapes under consideration. The uniqueness of $\mathbb{R}_p$ thus avoids possible errors related to the choice of either of them as a reference.



Histograms at $\Theta = 3^\circ$ resolution:

bolean $\mathcal{H}(\mathbb{X}, \mathbb{R}_p)$ = `[1 1 1 1 1 0 0 0 ... 0]`$_{1\times120}$
metric $\mathcal{H}(\mathbb{X}, \mathbb{R}_p)$ = `[0.14 0.21 0.24 0.28 0.12 0 0 0 ... 0]`$_{1\times120}$

Fig. 3. Computing spatial relations using radial line rotation ↻.

As depicted in Fig. 3, the line rotates over a cycle, and intersecting with object $\mathbb{X}$ ($\mathbb{A}$ or $\mathbb{B}$), generates a boolean histogram $\mathcal{H}$,

$$\mathcal{H}(\mathbb{X}, \mathbb{R}_p) = [I(\mathbb{R}_p, j\Theta)]_{j=0,\ldots,m-1}, \text{ where } I(\mathbb{R}_p, \theta_i) = \begin{cases} 1 \text{ if } line(\mathbb{R}_p, \theta_i) \cap \mathbb{X} \neq \emptyset \\ 0 \text{ otherwise.} \end{cases}$$

This is extended *wlog* to the sector defined by two successive angle values (θ_i and θ_{i+1}) and is normalised with respect to the total area of the studied object such that $\sum \mathcal{H}(.) = 1$. This histogram provides both spatial and structural information. Fig. 3 provides an illustration for an arbitrary object $\mathbb{X}$.

The histogram can be made rotationally invariant by projecting the smallest angle made by $\mathbb{R}_p$ and $\mathbb{C}_\mathbb{X}$ on the horizontal axis. Also, translation does not affect at all, since it uses $\mathbb{C}_\mathbb{X}$. Scaling does not produce any difference in $\mathcal{H}$ as it is normalised.

Vertex Signature via Shape Features. Each vertex has a distinct vocabulary type containing different shape and size information. Since spatial relations only encode relative positioning and point distributions, and do not completely exploit global shape information in the way shape descriptors do, we study the pertinence of $\mathcal{R}-$signature [19], region based Zernike moments (ZM) [20], generic fourier descriptors (GFD) [21] and shape context (SC) [22].

3 Symbol Recognition

Based on our symbol description, matching of two symbols is done by matching their corresponding ARGs. Consider two ARGs, query $G^q = (V^q, E^q, F_A^q, F_E^q)$ and database $G^d = (V^d, E^d, F_A^d, F_E^d)$, where the set of vertices $V = \{\mathbb{T}_1, \ldots, \mathbb{T}_t\}$, and the set of edges $E = \{E_1, \ldots, E_r\}$.

Our matching is straightforward i.e., matching has been made between the candidates only having the exact same set of vertices as well as exact labels. To generalise this, we define a binary indicator function $\tau_A^V : \Sigma_\mathbb{T} \rightarrow \{0, 1\}$ to check the presence of vertices in the ARG, where the value of $\tau_A^V(\mathbb{T})$ is 1 if $\mathbb{T}$ is present in V and 0, otherwise. For example, for the symbol shown in Fig. 2, the indicator $\tau_A^V = [0, 1, 1, 1]$ refers to the absence of *thick* components and the presence of *circle*, *corner* and *extremity*. Now, we can then set up bijective

matching functions: $\sigma : V^q \rightarrow V^d$ and $\varphi : E^q \rightarrow E^d$, respectively for vertices and edges. The fusion of both alignments provides the distance between two matched graphs G^q and G^d,

$$\text{Dist.}(G^q, G^d) = \alpha \sum_{t \in V} \delta(F_A^q(t), F_A^d(\sigma(t))) + (1 - \alpha) \sum_{r \in E} \delta(F_E^q(r), F_E^d(\varphi(r))),$$

where $\alpha \in [0, 1]$ and $\delta(a, b) = \sum_{l=1}^{L} ||a_l - b_l||_2$. The parameter α provides weight while matching.

- $\alpha = 0$: only vertex signature;
- $\alpha = 1$: only edge signature; and
- $\alpha = 0.5$: equal weights to both vertex and edge signature.

In our experiments, we provide equal weights for both relations and shape distribution of the whenever they are integrated together.

4 Experiments

4.1 Dataset, Ground-truth and Evaluation Metric

Fig. 1 gives an overview of the dataset we are using for our experiments. The global dataset is composed of roughly 500 different known symbols, some of which come from [5]. It shows that symbols may either be very similar in shape – and only differ by slight details – or either be completely different from a visual point of view. Symbols may also be composed of other known and significant symbols and need not necessary be connected.

Since there is no absolute ground-truth associated to our dataset, we have asked 6 volunteers to manually select what they consider as "similar" symbols, and we have merged their inputs to reduce possible subjective bias. They have chosen the candidates which have similar visual overall appearance or contain significantly similar parts with respect to the chosen query. In our testing protocol, we consider that a result returned from an algorithm is correct if at least one evaluator has selected the same result among similar items. In more formal terms, for each query the "ground-truth" is considered to be the set of symbols formed by the union of all human selected sets.

For every query, we rank the symbols at the output based on distance measure as described in Section 3. Since the number of similar symbols, according to the ground-truths, may vary a lot from one query to another, we use retrieval efficiency [23] as a measure for retrieval quality. For a chosen query and for a fixed number of K returned symbols, it can be expressed as,

$$\eta_K = \begin{cases} n/N \text{ if } N \leq K \\ n/K \text{ otherwise,} \end{cases}$$

where n is the number of returned relevant symbols and N the total number of relevant symbols.

4.2 Results and Discussions

We perform a series of tests, focussing on three major issues one after another.

1. Choice of optimal resolution for radial line model (as edge signature) and comparison with other relation models;
2. Choice of the best shape descriptors on vocabulary (as vertex signature) and comparison with global symbol shape descriptors;
3. Integration of both vertex and edge signatures.

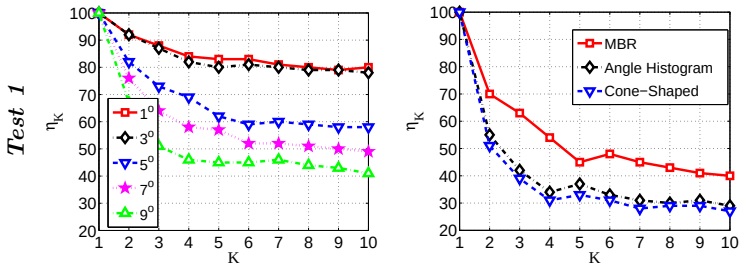
In what follows, we compute the average retrieval efficiency for 30 queries, and vary K from 1 to 10.

Test 1. We consider the influence of different resolutions Θ in our edge signature. Its value represents the trade-off between the optimal choice of resolution – and thus precision of spatio-structural information capture – and time/space requirements. Without surprise, the lower Θ, the better the results. Following results in Fig. 4 (a), and given the relatively low gain of efficiency between 3° and 1°, we adopt 3° for the rest of our experiments. We then compare our approach with state-of-the-art spatial relation models: cone-shaped [15], angle histogram [17] and MBR [16] as shown in Fig. 4 (b). Compared to the best performing model (MBR), our radial-line approach increases performance with a substantial difference. These results were already reported in [13].

Test 2. In section 2 we already listed the set of global shape-descriptors we consider state-of-the-art: $\mathcal{R}-$signature [19], ZM [20], GFD [21] and SC [22]. For those descriptors, it is important to fit the best parameters. For GFD, we have tuned the radial and angular frequency parameters to achieve the best performance. For SC, we attempt to follow the indications given in [22] but they are restricted by the number of sample points of some symbols (i.e., we have images ranging from a few tens of pixels to thousands of pixels). In the case of ZM, we have used 36 *Zernike* functions of order less than or equal to 7. For Radon, projecting range is $[0, \pi[$.

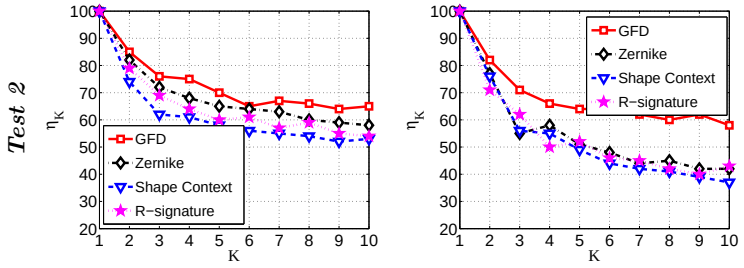
We first employ them as vertex signatures only and then confront them with the same shape descriptors, applied to the overall shape. This comparison is illustrated in Fig. 5 (a) and (b), respectively. In this test, SC, $\mathcal{R}-$signature and ZM are lagging behind GFD. Therefore, we consider GFD to be the best performing descriptor and will use it as a benchmark for further experiments.

Test 3. We integrate both edge and vertex signatures in the ARG described in Section 2. However, integrating signatures on all vertices is not necessary since our spatial relation signature already carries sufficient information for those vocabulary types less influenced by shape variations (*corner* and *extremity*, for instance). Therefore, we have examined the use of signature from some specific types only, and on some combinations like, *thick*, *thick* and *circle*, *thick*, *circle* and *corner*, and so on. It is important to notice that we have never left out *thick* components from all tested combinations because those vocabulary elements



(a) Radial line: $\Theta = \{1°, \ldots, 9°\}$ (b) Basic relation models

Fig. 4. Average retrieval efficiency using spatial relations.



(a) Vertex signature (b) Overall symbol shape[1]

Fig. 5. Average retrieval efficiency using global signal based shape descriptors.

provide the largest amplitude in shape and size variations. To handle this, we use all aforementioned global signal based descriptors as vertex signatures in order to see how well they improve retrieval efficiency. In our tests, substantial advancement is achieved from the combination of vertices, labelled with *thick* and *circle* vocabulary types. Fig. 6 shows results from using different shape signatures combined with the edge signature using a radial line model at 3^o resolution. GFD provides the best results. Surprisingly, not all vertex shape descriptors improve upon the plain edge signatures, but some even decrease in performance. $\mathcal{R}-$signature provides an example.

Comparison. In Fig. 7, a comparison is made among the best of all experiments: MBR from the basic spatial relation model and GFD from shape descriptors. Our method outperforms all with a difference of more than 16%. For a few queries, an illustration is shown in Fig. 8.

Another important issue accounting for retrieval performance is execution time. In Table 1, we provide it based on comparison of the methods in Fig. 7. This proves that the increase in recogition efficiency comes at a cost in execution time. This cost remains in a reasonable limit, however.

[1] Compared to our previous work [13] where the $1D$ (vanilla) version of the $\mathcal{R}-$signature has been employed, the complete version reported here provides better performance, without altering the global validity of our findings, however.

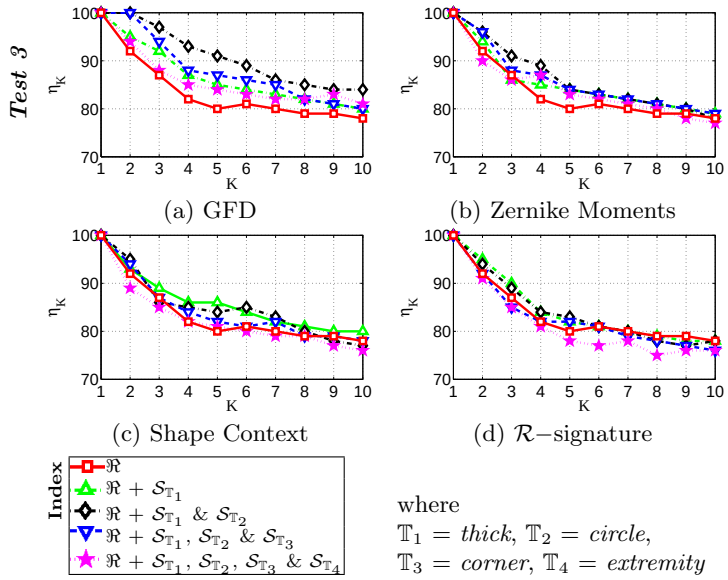


Fig. 6. Average retrieval efficiency while integrating vertex signature in our ARG framework.

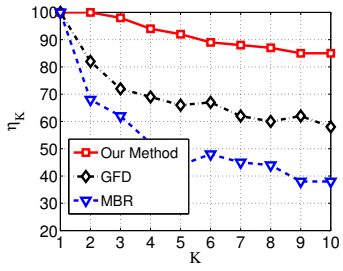


Fig. 7. Comparison among the best of all categories.

5 Conclusions

We have presented an ARG based symbol description method, where relational signatures formalise all possible connections between the vocabulary types which are labelled with global shape signatures. Our method has proven to significantly outperform state-of-the-art basic spatial relation models as well as global signal based descriptors.

References

1. Cordella, L.P., Vento, M.: Symbol recognition in documents: a collection of techniques? IJDAR **3**(2) (2000) 73–88
2. Lladós, J., Valveny, E., Sánchez, G., Martí, E.: Symbol Recognition: Current Advances and Perspectives. In: GREC – Algorithms and Applications. Volume 2390. Springer Verlag (2002) 104–127

	Test 1		Test 2	Test 3
	MBR	Edge sign.	GFD	Edge sign. + GFD vertex sign.
Time	02	04	09	14

Table 1. Average running time (sec.) for matching for a single pair.

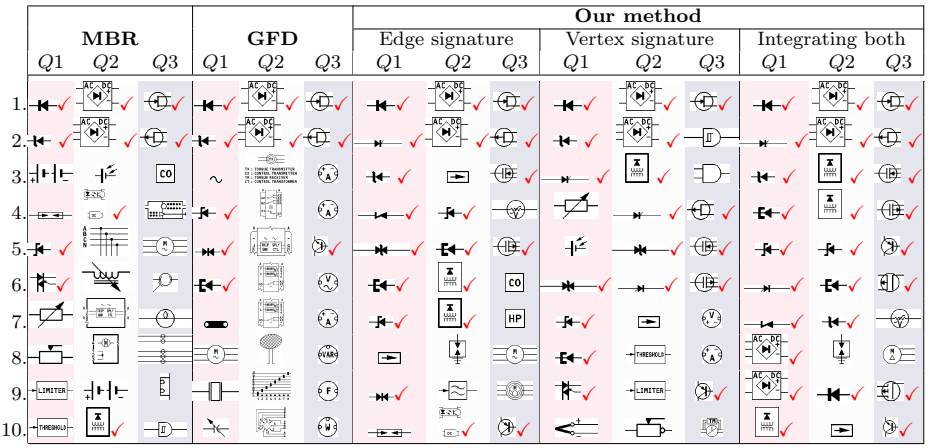


Fig. 8. Visual illustration of symbol ranking at the output for a few queries: $Q1$, $Q2$ and $Q3$, showing ✓ for true retrieval and false, otherwise. The first symbol on the top always corresponds to the chosen query. Symbols are ranked from top to bottom based on decreasing order of similarity.

3. Tombre, K., Lamiroy, B.: Pattern recognition methods for querying and browsing technical documentation. In: CIARP. (2008) 504–518
4. Santosh, K.C., Lamiroy, B., Ropers, J.P.: Inductive logic programming for symbol recognition. In: ICDAR. (2009) 1330–1334
5. Tooley, M., Wyatt, D.: Aircraft electrical and electronic systems: principles, operation and maintenance. Aircraft engineering principles and practice. Butterworth-Heinemann (2008)
6. Zhang, D., Lu, G.: Review of shape representation and description techniques. PR **37**(1) (2004) 1–19
7. Tombre, K., Ah-Soon, C., Dosch, P., Habed, A., Masini, G.: Stable, robust and off-the-shelf methods for graphics recognition. Volume 1. (1998) 406
8. Barrat, S., Tabbone, S.: A bayesian network for combining descriptors: application to symbol recognition. IJDAR **13**(1) (2010) 65–75
9. Terrades, O.R., Valveny, E., Tabbone, S.: Optimal classifier fusion in a non-bayesian probabilistic framework. IEEE PAMI **31**(9) (2009) 1630–1644
10. Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. IJPRAI **18**(3) (2004) 265–298
11. Lladós, J., Martí, E., Villanueva, J.J.: Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. IEEE PAMI **23**(10) (2001) 1137–1143
12. Valveny, E., Martí, E.: A model for image generation and symbol recognition through the deformation of lineal shapes. PRL **24**(15) (2003) 2857–2867

13. Santosh, K.C., Lamiroy, B., Wendling, L.: Symbol recognition using spatial relations. PRL **33**(3) (2011) 331–341
14. Rendek, J., Masini, G., Dosch, P., Tombre, K.: The search for genericity in graphics recognition applications: Design issues of the qgar software system. In: DAS. Volume 3163. 366–377
15. Miyajima, K., Ralescu, A.: Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations. Fuzzy Sets and Syst. **2**(65) (1994) 225–236
16. Papadias, D., Theodoridis, Y.: Spatial relations, minimum bounding rectangles, and spatial data structures. Int. J. Geographical Inform. Sci. **11**(2) (1997) 111–138
17. Wang, X., Keller, J.: Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches. Fuzzy Sets and Syst. **101** (1999) 5–20
18. Freeman, J.: The modelling of spatial relations. CGIP **4** (1975) 156–171
19. Tabbone, S., Wendling, L., Salmon, J.P.: A new shape descriptor defined on the radon transform. CVIU **102**(1) (2006) 42–51
20. Kim, W.Y., Kim, Y.S.: A region-based shape descriptor using zernike moments. Signal Process.: Image Commun. **16**(1-2) (2000) 95–102
21. Zhang, D., Lu, G.: Shape-based image retrieval using generic fourier descriptor. Signal Process.: Image Commun. **17** (2002) 825–848
22. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE PAMI **24**(4) (2002) 509–522
23. Kankanhalli, M.S., Mehtre, B.M., Wu, J.K.: Cluster-based color matching for image retrieval. PR **29** (1995) 701–708